

科学研究費助成事業 研究成果報告書

平成 29 年 5 月 3 日現在

機関番号：32634

研究種目：若手研究(B)

研究期間：2014～2016

課題番号：26750114

研究課題名（和文）ノンパラメトリック推定に基づくテスト理論モデルの研究

研究課題名（英文）Study on Test Theory based on Nonparametric Estimation

研究代表者

高野 祐一（Takano, Yuichi）

専修大学・ネットワーク情報学部・准教授

研究者番号：40602959

交付決定額（研究期間全体）：（直接経費） 1,600,000円

研究成果の概要（和文）：項目反応理論とは、被験者の能力を正確に測定可能なテストを作成し、運用するための現代的なテスト理論である。本研究では、「被験者の能力」と「設問の正答確率」との関係を表す項目特性曲線を、特定の関数形を仮定することなく推定するノンパラメトリック項目反応理論を対象としている。本研究では、既存モデルに対する混合整数最適化問題による定式化や、項目特性曲線の過剰適合を軽減する平滑化制約を提案し、提案モデルの推定アルゴリズムを作成した。また既存の項目反応理論モデルとの比較実験により、提案モデルの有効性を検証した。

研究成果の概要（英文）：Item response theory is a modern test theory for the design, analysis, and scoring of tests. This research is concerned with the nonparametric item response theory (NIRT) for estimating various forms of item characteristic curves (ICCs), which express the probability of a correct answer as a function of the latent abilities of the examinees. This research derived mixed-integer optimization formulations for existing NIRT models, proposed smoothness constraints for preventing overfitting of nonparametric ICCs, and developed algorithms for estimating these NIRT models. Computational results demonstrated the effectiveness of the NIRT models in comparison to existing IRT models.

研究分野：数理最適化

キーワード：テスト理論 数理最適化 機械学習 統計計算 アルゴリズム

1. 研究開始当初の背景

テスト理論とは、被験者の能力を正確に測定可能なテストを作成し、運用するための理論である。現代的なテスト理論である項目反応理論では、各設問に対して「被験者の潜在的な能力」と「設問の正答確率」との関係を表す項目特性曲線を推定する。項目反応理論では、この項目特性曲線を利用して、被験者のテスト結果から被験者の能力を最尤法によって推定する。

広く使われているパラメトリック項目反応理論では、項目特性曲線の形状としてロジスティック曲線などの特定の関数形を仮定する。一方で本研究では、項目特性曲線の形状として特定の関数形を仮定せず、各能力に対応する正答確率を直接求めるノンパラメトリック項目反応理論に着目する。ノンパラメトリック項目反応理論では、得られたデータに応じて多様な項目特性曲線を推定することが可能となる。

しかし、ノンパラメトリック項目反応理論には二つの欠点がある。第一の欠点として、項目特性曲線の自由度が高過ぎるために過剰適合が起こるといった問題が指摘されている。被験者のテスト結果は、当て推量やケアレミスなどの誤差を含んで観測される。よって、このような誤差を含むデータに過剰に適合してしまつと、正しい項目特性曲線を推定することができない。第二の欠点として、ノンパラメトリック項目反応理論では、被験者のテスト結果から「各設問の項目特性曲線」と「各被験者の能力」を同時に推定するための手法が確立されていない。

2. 研究の目的

本研究の目的は、ノンパラメトリック項目反応理論の欠点を克服し、実用化のレベルまで性能を高めることである。具体的には、ノンパラメトリック項目反応理論に関して以下の課題に取り組む。

(1) 過剰適合により推定精度が悪化するという問題点を解決する。

(2) 「各設問の項目特性曲線」と「各被験者の能力」を同時に推定するための手法を開発する。

3. 研究の方法

(1) 正則化(罰則付き推定)とロバスト最適化の手法を利用して、過剰適合を制御可能なノンパラメトリック項目反応理論モデルを開発する。

(2) 数値最適化手法を利用して、ノンパラメトリック項目反応理論の「各設問の項目特性曲線」と「各被験者の能力」を同時に推定するための実用的なアルゴリズムを開発する。

(3) 実際のテスト結果のデータや、仮想的な

テストデータを利用して、ノンパラメトリック項目反応理論の実用性と推定精度を検証する。

4. 研究成果

(1) ノンパラメトリック項目反応理論の代表的なモデルとして、単調等質性モデルと二重単調性モデルがある。単調等質性モデルでは、項目特性曲線の単調増加性(能力の高い被験者ほど正答確率が高いこと)が制約として課される。二重単調性モデルでは、さらに各設問の項目特性曲線が交差しない(被験者の能力によって設問の難易度が逆転しない)ことが制約として課される。本研究では、単調等質性モデルと二重単調性モデルに対して、混合整数最適化問題による定式化を提案した。この定式化では、要求される制約条件の下で「項目特性曲線」と「被験者の能力」を同時に決定するが、混合整数最適化問題を厳密に解くことは非常に難しい。そこで本研究では、良質な解を効率的に求めるための方法として、「項目特性曲線」と「被験者の能力」を交互に推定するアルゴリズムを作成した。数値実験の結果、多数の設問において被験者の能力が多峰型分布に従う場合に、一般的な2母数ロジスティック項目反応理論モデルと比較して、提案手法は推定精度が優れていることを検証した。

(2) 上述の単調等質性モデルでは、項目特性曲線の単調増加性が制約として課されるが、この制約だけだと依然として項目特性曲線の自由度が高く、過剰適合が生じて推定精度が低くなることがある。一方で二重単調性モデルでは、さらに各設問の項目特性曲線が交差しないことが制約として課されるが、実際の項目特性曲線は交差することが多く、制約としては強過ぎると言える。そこで本研究では、単調等質性モデルに項目特性曲線の平滑化制約を加えたノンパラメトリック項目反応理論モデルを提案した。本研究で提案する平滑化制約は、項目特性曲線をロジスティック曲線に近付けることで項目特性曲線の自由度を抑える制約となっており、この意味で提案モデルはパラメトリックモデルとノンパラメトリックモデルの両方の利点を活かしたモデルと言える。さらに本研究では、効率的にモデルを推定するために、EMアルゴリズムに基づく推定手法を実装した。数値実験の結果、被験者の能力が多峰型分布に従う設問がある場合には、一般的な2母数ロジスティック項目反応理論モデルと比較して、提案モデルは推定精度が優れていることを検証した。さらに、提案モデルは単調等質性モデルよりも推定精度が優れていることも実証した。

(3) 関連研究として、インターネット上の商品販売サイトにおける利用者の商品選択確率を推定するための最適化モデルを提案し

た。提案モデルでは、利用者の各商品に対するページ閲覧の最新度と頻度に着目し、最新度と頻度の組に対して商品選択確率を推定する。さらに提案モデルでは、商品選択確率の形状に関して単調性・凸性・凹性などの制約を課すことで、データ数が少ない場合でも過剰適合による推定精度の悪化を軽減することができる。数値実験の結果、ロジスティック回帰やサポートベクター分類などの既存手法よりも、提案モデルの推定精度が優れていることを検証した。

(4) 関連研究として、ロジスティック回帰モデルの変数選択問題に対する混合整数最適化モデルを提案した。ロジスティック回帰モデルの変数選択問題は、対数尤度関数が非線形関数となるために扱い難い。そこで本研究では、目的関数に現れるロジスティック損失関数に対して区分線形近似を施すことで、混合整数線形最適化問題に帰着し、汎用の整数最適化ソフトウェアによる求解を可能とした。数値実験の結果、ステップワイズ法や罰則付き回帰などの手法と比較して、提案モデルは情報量規準値の良い変数集合を選択できることを検証した。本研究で提案した区分線形近似手法は、逐次ロジット回帰による順序付き多クラス分類にも拡張することが可能であり、既存の二次近似手法よりも良い推定結果が得られることを示した。

(5) 関連研究として、重回帰モデルの説明変数間の階層構造を考慮した混合整数最適化による変数選択手法を提案し、対象店舗における購買点数に着目した店舗選択要因の分析を行なった。提案手法は、商品分類間の階層構造を活用してモデル構築の信頼性を高めることができ、計算効率が向上するという利点もある。数値実験では、ステップワイズ法や罰則付き回帰と性能を比較し、提案手法の有効性を検証した。また、提案手法を用いて3つの業態に対する消費者の店舗選択要因を分析した結果、コンビニエンスストアは利便性、ドラッグストアは価格や商品構成によって選択されていること、食品スーパーは家族を持つ女性に支持されており、健康的な食品が主な選択要因となっていることを示した。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計4件)

Toshiki Sato, Yuichi Takano, Ryuhei Miyashiro, Piecewise-Linear Approximation for Feature Subset Selection in a Sequential Logit Model, Journal of the Operations Research Society of Japan, 査読有, Vol.60, No.1, pp.1-14 (2017).

DOI: 10.15807/jorsj.60.1

Toshiki Sato, Yuichi Takano, Ryuhei Miyashiro, Akiko Yoshise, Feature Subset Selection for Logistic Regression via Mixed Integer Optimization, Computational Optimization and Applications, 査読有, Vol.64, No.3, pp.865-880 (2016).
DOI: 10.1007/s10589-016-9832-2

Jiro Iwanaga, Naoki Nishimura, Noriyoshi Sukegawa, Yuichi Takano, Estimating Product-Choice Probabilities from Recency and Frequency of Page Views, Knowledge-Based Systems, 査読有, Vol.99, No.1, pp.157-167 (2016).
DOI: 10.1016/j.knosys.2016.02.006

Yuichi Takano, Shintaro Tsunoda, Masaaki Muraki, Mathematical Optimization Models for Nonparametric Item Response Theory, Information Science and Applied Mathematics, 査読無, Vol.23, pp.1-18 (2016).
<http://id.nii.ac.jp/1015/00009881/>

[学会発表](計4件)

佐藤俊樹, 高野祐一, 中原孝信, 混合整数最適化による階層的変数選択を用いた店舗選択要因の分析, 第15回情報科学技術フォーラム(FIT 2016), 2016年9月7日~9月9日, 富山大学(富山県富山市).

Yuichi Takano, Toshiki Sato, Ryuhei Miyashiro, Akiko Yoshise, Feature Subset Selection for Logistic Regression via Mixed Integer Optimization, Workshop on Advances in Optimization, 2016年8月12日~8月13日, TKP Shinagawa Conference Center(東京都港区).

Yuichi Takano, Toshiki Sato, Ryuhei Miyashiro, Akiko Yoshise, Feature Subset Selection for Logistic Regression via Mixed Integer Optimization, The 28th European Conference on Operational Research (EURO 2016), 2016年7月3日~7月6日, Poznan University of Technology, Poznan (Poland).

Yuichi Takano, Feature Subset Selection for Linear/Logistic Regression via Mixed Integer Optimization, The 22nd International

Symposium on Mathematical Programming
(ISMP 2015), 2015年7月12日~7月17
日, Wyndham Grand Pittsburgh Downtown,
Pittsburgh (U.S.A.).

6. 研究組織

(1) 研究代表者

高野 祐一 (TAKANO, Yuichi)

専修大学・ネットワーク情報学部・准教授

研究者番号: 40602959